The Sonification and Learning of Human Motion

Kevin M. Smith

California State University, Channel Islands
One University Drive,
Camarillo, California 93012
k2msmith@gmail.com

David Claveau

California State University, Channel Islands
One University Drive,
Camarillo, California 93012
david.claveau@csuci.edu

ABSTRACT

This paper examines how sonification can be used to help a student emulate the complex motion of a teacher with increasing spatial and temporal accuracy. The system captures a teacher's motion in real-time and generates a 3-D motion path, which is recorded along with a reference sound. A student then attempts to perform the motion and thus recreate the teacher's reference sound. The student's synthesized sound will dynamically approach the teacher's sound as the student's movement becomes more accurate. Several types of sound mappings which simultaneously represent time and space deviations are explored. For the experimental platform, a novel system that uses low-cost camera-based motion capture hardware and open source software has been developed. This work can be applied to diverse areas such as rehabilitation and physiotherapy, performance arts and aiding the visually impaired.

1. INTRODUCTION

In this paper we explore the use of sonification as a feedback mechanism for learning a complex 3-D motion trajectory in real-time. Most techniques for learning movement, whether they are used for dance, sports or other articulated motion are visual. These can include interactive methods such as live demonstration and/or video recording of the movement for playback and coaching. These methods are predominantly visual with assisted verbal instruction provided by a teacher or coach. Here we focus on a different approach – the augmentation of the learning process with layered *sound* to provide a sonic feedback mechanism for learning precise motion.

With the advent of relatively new and inexpensive technologies such as the Microsoft Kinect (2010) [1] and Leap Motion Leap (2013) [2], it is now possible to capture elements of human motion in 3-D and in real-time on relatively low-cost hardware. With the fast evolution of this hardware and the consumer demand for games and applications that use them, we expect full-body skeletal capture capabilities to evolve rapidly in terms of performance and accuracy. Inspired by these new developments, we specifically explore the idea of using sonification as an auditory feedback mechanism for learning how to reproduce a 3-D motion path of a hand or an endpoint of a limb. In particular, we are interested in looking at how layered sound can be used to provide *concurrent* feedback on both *spatial accuracy and timing* of the motion. It is our goal that once we solve the atomic problem of a single motion path, we can ultimately extend this research to more complex hierarchical motion containing potentially many motion paths of multiple joints and multiple bodies.

Existing work in the area of aiding movement by the use of sonification focuses on a number of different topics. In the work of Rober and Masuch[3], the use of interactive auditory environments and 3-D sound rendering to explore virtual auditory environments is the focus in the design of a framework. Effenberg[4] used sonification to assist in the reproduction of human movements, showing that sound can provide additional information in the accurate reproduction of jumps and other athletic movements. In the work of Kleiman-Weiner and Berger[5], arm swinging motion, using the example of the golf swing, is sonified. Other work includes the use of sound for physiotherapy. Feedback is an area studied by Pauletto and Hunt[6]. In their work, the sonification of EMG signals gathered in a clinical environment provide auditory display to the therapist in real-time, producing sound with muscle movement that is audible in the room when visual displays are not always within view[7]. PhysioSonic[8] was developed as a system to map motion capture data to sound to provide auditory feedback for physiotherapy and training.

In addition to feedback, there are existing projects using real-time articulated motion for *generating* sound and music. In these interactive performances, performance gestures are translated to music and motion graphics, allowing the body to generate sound and visual effects. *Synapse*[9] and *The V Motion Project*[10] are two such example projects which both use the *Kinect* device for capturing the motion.

Building on this existing work we specifically look at sonification feedback for learning precise motion along a path with spatial and timing accuracy as measured by the distance between an endpoint and the target reference path. We will focus on the development of a portable laptop system that uses low-cost consumer capture devices. In the first section of this paper we describe the overall system design which includes the internal data representation of the 3-D motion path with timing, sound mappings and synthesis. This will be followed by a description of the actual implementation of our experimental system, which we call *SoundTracer*. Following that, the implementation section will look at some of the initial results of using our system and finally present conclusions and opportunities for future work.

2. SYSTEM DESIGN

To best describe the design and operation of the system it is good to start with a use-case which will describe the basic goals and interactions between the system and the actors who will use the system. Then we will discuss the workflow of the system in more detail. Figure 1 shows a schematic representation of the workflow.

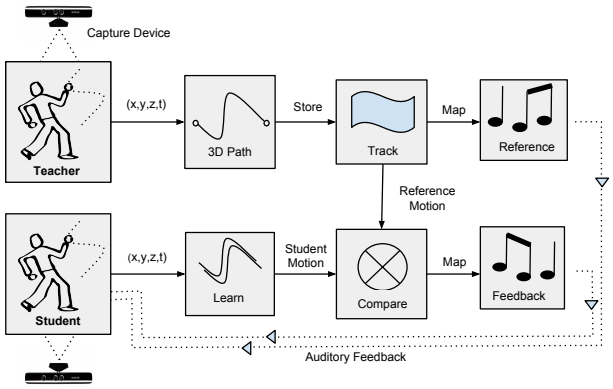

Figure 1 – System Workflow: difference comparison between sonification of the teacher's motion and the student's.

The *teacher* creates a movement that is captured and recorded in 3-D space in real-time. Any major joint in the skeleton (including the torso) can be tracked, provided it is configured in advance. The *student* then has the goal of learning the motion created by the teacher by performing the motion as accurately as possible. By *performing* we mean the ability to reproduce the motion of the teacher as accurately as possible with respect to achieving the same trajectory of motion in both space and time.

The principal goal of the system is to use sound as a feedback mechanism to assist the student in learning the motion. When the teacher creates the motion, *sound* is automatically generated for the reference motion and stored. When the student attempts to reproduce the motion sound is also generated. In order to reproduce the motion accurately, the student must also reproduce the sound that was generated by teacher. Any deviation in the motion will produce a corresponding deviation in the sound generated by the student. How the sound is generated or mapped from the 3-D motion path over time is described in Section 2.2.

2.1. Data Flow

The student or teacher's motion is captured in real-time from a 3-D tracking device. At a capture rate of 30Hz, we are able to obtain a stream of sequential 3-D points for the capture of one joint in the form of:

$$\mathbf{P} = (x, y, z, t)$$

Where **P** represents a single point in data stream at time *t*.

We represent the 3-D motion path by the stream of points which constitutes a piecewise linear approximation of a curve. From this curve we are able to easily lookup features such as the location on the path at any point in time, the distance to the nearest point on the curve from any point and the length along the curve at any point using a piecewise linear approximation. It is worth mentioning that we considered more mathematically complex (and computationally more expensive) spline representations of the motion path and we determined that given most devices were capable of returning points at 30 Hz or more, the density of the data was sufficient to approximate the paths for our purposes at the move duration we experimented with.

As shown in figure 1, a 3-D path generated by the teacher is stored in a *Track.* These are persistent objects, which can be stored on disk for later retrieval by the student when learning. The same mechanism for generating a motion path is used by both the teacher and student. Once the data is stored in the Track, it is processed, so that when the track is *played back,* automatically generated sound will accompany the motion. This sonification process will be discussed in more detail in the next section.

After a teacher records a path to be learned in a *Track*, the student can retrieve this path and initiate a learning session. The learning session enables real-time capture and comparison of the student's motion path with the teacher's. A comparison is made both spatially and temporally. The student has the option of obtaining feedback on each component independently or concurrently. For the spatial comparison, the distance between the student's current point (x, y, z) at any point in time is compared with the nearest point on the teacher's path. The distance between these two points determines the amount of spatial error present at any time. This error can be used as an input to the sonification of the student's motion. If there is no error (within a preset tolerance) the sound is not modified.

For the temporal comparison, the progress along the motion path by the student in terms of curve length is calculated at the current time, *t.* If the student is ahead of where the teacher should be (the curve length is longer), then the student is moving too fast and the difference in progress is propagated to the sonification process and the feedback sound can be modified. Conversely if the curve length is shorter, then the student is behind the teacher and moving too slowly and that difference is also propagated. If the progress along the path is the same as the teacher's (within a preset tolerance), the sound is not modified.

2.2. Sonification

In designing the auditory feedback for the learning process, we decided on two primary goals. The first goal was that we

wanted to provide a way to simultaneously allow the student to correct for both spatial accuracy and timing accuracy concurrently. This system should provide feedback that would enable the student to make real-time corrections to both the path of travel and the progress along that path with respect to time. For the timing aspect the system should be sensitive and provide feedback to changes in rate of progress (acceleration) and speed (velocity) over the trajectory. This may be challenging for the type of articulated motion present in many applications (performance arts, sports, rehabilitation, therapy) which can have a very wide range of space and time characteristics. A slow coordinated movement over several seconds would have different sonification requirements than a fast movement, such as a golf swing, which has a short duration. We look to design a mapping that will help with both slow and fast motion. A second goal is the aesthetic quality of the sound. With an objective of serving artists and performers, we want the sound to be as pleasing as possible.

2.2.1. The Attack, Decay, Sustain, Release Model

To incorporate timing and spatial information in auditory feedback for fast motion we use the concept of an *envelope.* By modeling the *attack, decay, sustain, release envelope (ADSR)* used in electronic instrument synthesis, we apply a control envelope as a multiplier to the overall output or volume of the sound being generated. In a real instrument such as a guitar, time is a linear quantity and the ADSR progresses from the time the instrument is plucked to the time the vibrating string stops moving. In our model, we generate an ADSR envelope to match the teacher's duration of the recorded motion. When the student attempts to reproduce the motion, their progress along the path with respect to time actuates the ADSR. As an example, using a simpler ADSR similar to a stringed instrument, if the student is initially too fast, the attack component of the sound will come sooner. If the student is late in the second part of the move, the sound will sustain longer.

ADSR incorporates timing information into the model, but we also need a way to provide corrective feedback for spatial variances. For this approach we use a simple pitch shifting method. When the student deviates from the prescribed path, the pitch will change proportionally to the distance from the location on the path where the student should be at that point in time. The combined layering of the ADSR as envelope for volume and the shifting of the pitch provide two concurrent degrees of freedom for auditory feedback.

The effect of the ADSR will depend on the type of envelope chosen and the source sound that it will operate on. Figure 2 shows two envelope examples that were used in our experiments. For our initial experiments, we found that a bell-shaped curve worked the best. Since the rise in the energy of the sound comes roughly half way through the motion, it was easier to learn the timing of the motion based on mental correlation on where the midpoint of the sound should be with respect to the midpoint of the motion.

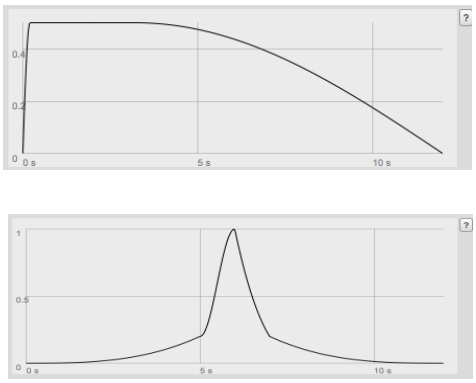

Figure 2 - Envelope Curves Used. Top: a ADSR with early attack and longer sustain. Bottom: An ADSR with a bell-shaped curve

For the source sound to be processed by the envelope, we required a sound that (1) would sound aesthetically pleasing and (2) could be sustained for longer periods of time. The instrument that we tried first was the flute. For our tests we used a synthesized flute sound using waveguide methods [11][12]. Our model included parametric control over breathing, vibrato and pitch with the addition of the ADSR curve to control overall energy of the output sound over time. The student would not only be reproducing the pitch of the flute, but the overall volume envelope during the course of the motion.

2.2.2. Sampled Music Feedback

With ADSR we explored an approach that would enable us to layer spatial and temporal information to provide auditory feedback for fast movement. In order to explore the approach of using a pre-recorded musical *sound track* for providing auditory feedback, particularly for slower movement that may be more performance-oriented such as ballet dancing, we also researched a *sampled music* approach. In this approach (rather than using a synthesized single instrument with an ADSR), the teacher performs the move to sampled music. The goal of the student is to reproduce the move and the music at the same time. Any deviations to the motion path will cause a pitch shift in the music. Any deviation to the timing along the path will cause the tempo of the music to change. This method, again, provides us with two degrees of freedom simultaneously for auditory feedback on spatial and temporal accuracy.

In the next section, we discuss the full implementation of Sound*Tracer,* an application we developed to experiment with these ideas.

3. IMPLEMENTATION

We have developed a system called Sound*Tracer* to enable a student to practice real-time reproduction of motion paths

created by a teacher. A block diagram of the system is shown in figure 3.

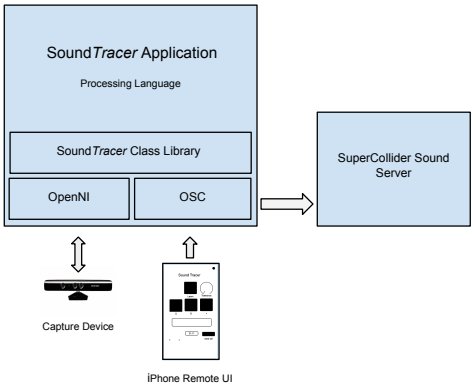

Figure 3 - Sound*Tracer* system consisting of Processing application communicating with Capture Device (Kinect), OSC control devices and the SuperCollider Sound Server.

Sound*Tracer* is based purely on open-source software components. We used *Processing*[13] as the implementation language because of its strength as a rapid prototyping language and the out-of-the-box experience enabling us to get it installed and running quickly with good library support for communicating with other devices and servers.

Although the code is generic enough to accept data from any 3-D tracking device, we used the first generation *Microsoft Kinect* device along with public domain libraries for the *OpenNI*[14] driver. The *Kinect* is limited in resolution and tracking stability in comparison to a professional optical motion capture system, however, the cost and convenience of this device made it practical for testing in most environments. In addition to the *Kinect*, we have also done some limited testing with the *Leap Motion* device, which provides smaller scale (but more precise) tracking for finger motion. This could be interesting for learning smaller body movements involving hand gestures (See Section 5.)

The generation of synthesized sound for audio is a very mature field. Rather than implement our own software synthesizer, we chose to use the SuperCollider[15] language and server. The flute sound, which was discussed in Section 2.2.1, was generated using a waveguide flute example *SynthDef* in the *SuperCollider* language. The ADSR envelope was implemented in the same language using an envelope generator which can poll values by index (for more information see *IEnvGen* in the *SuperCollider Reference)* [16].

In order for Sound*Tracer* to communicate with SuperCollider, we use an open-source Processing library (oscP5) [17] which supports the Open Sound Control Protocol (OSC) [18]. The dual benefit of this is (1) the SuperCollider server uses OSC as it's native protocol to receive messages from it's language module and (2) we can use OSC as a channel for communicating with Sound*Tracer* using other mobile devices such as the iPhone, which has a number of customizable OSC-based apps available. In this implementation, we built an iPhone custom user-interface on TouchOSC [19], which can control Sound*Tracer* remotely making a very convenient interface (and almost a necessity) while motion testing.

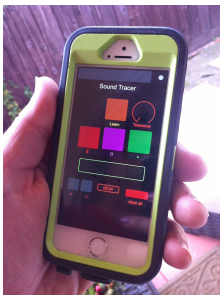

Figure 4 - iPhone interface for Sound*Tracer*

As discussed in Section 2, the purpose of Sound*Tracer* is to provide features for the real-time capture of 3-D motion, the storage of this data in the form of Tracks, sonification of the 3-D motion data and the real-time comparison of this data with student attempts to recreate the motion. In the picture sequences figures 5 and 6 we show examples illustrating a sample motion path and how spatial (top) and temporal testing (bottom) can be accomplished. Both can be done individually or concurrently. Visual aids for the motion path, target point, distance to target and color coding for the student's path to indicate on-target proximity are provided in addition to the sonification for feedback.

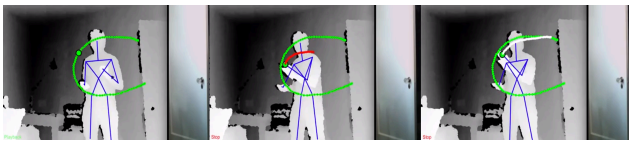

Figure 5 - Sound*Tracer* Shot Sequence for Spatial Testing. Reference motion (green) created by teacher in first frame. Middle: Out of bounds (red) motion path; Last: Motion correctly approaching (white) reference motion within tolerance.

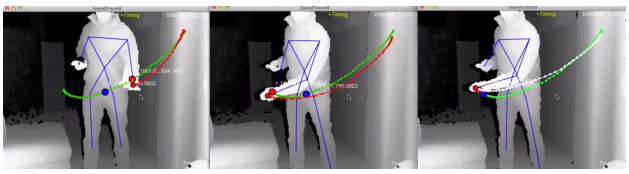

Figure 6 - Sound*Tracer* Shot Sequence for Temporal Testing. In first frame, student is early and behind target point. Middle: student is too fast and arrives at end of path ahead of target point. Last Frame: student is on track to arrive at the same time as target.

4. SYSTEM TESTING

To date we have conducted preliminary tests on the system with several users using the ADSR sonification method as described in Section 2.2. The goal of the testing is to provide initial feedback on the workflow process in order to increase usability of the system and to provide subjective feedback from the testers. Once this milestone has been achieved we can form the basis for a more rigorous study in the future.

Three levels of sonification were tested: (1) Visual aids only were presented to the user. No sonification of the 3-D data was used (2) Sonification + Visual. Sonification was used as an aid in addition the visual aids provided and (3) Sonification only. Sound was used as the main feedback mechanism. Minimal visual cues were provides. Note that in the last case, we still maintained some minimal visual cues to show the user where the motion path start/end is in order to provide a gate for measurement. In the case where visual aids were used, we provided a display of the full motion path with target tracking markers on the path to show the student's current position with respect to the teacher's. The example data in figure 7 shows a typical number of trials for a student for the Sonification + Visual case (2).

User (n)	Trial (n)	Mapping (ADSR, SAMPLED)	Level (VIS, VIS+SON, SON)	Planar Horiz	Planar Vert	Mixed 3D
1	1	ADSR	VIS+SON	112	194	407
	2	ADSR	VIS+SON	68.7	55	227
	3	ADSR	VIS+SON	98.8	85	153
	4	ADSR	VIS+SON	82.7	47	109
	5	ADSR	VIS+SON	59.1	77	148
2	1	ADSR	VIS+SON	147	74.6	398
	2	ADSR	VIS+SON	81.2	48.9	294
	3	ADSR	VIS+SON	55.1	54.36	188
	4	ADSR	VIS+SON	77	76.46	146
	5	ADSR	VIS+SON	52.1	51.47	127

Figure 7 – Example Sonification data for a test user (student)

For each trial, three moves were scored, one with planar motion in a horizontal plane, a second with vertical planar motion and, finally, a more complex full 3-D motion path with movement in all axial directions. We used a *root mean square* technique to score the difference between the student's curve the teacher's curve (at each time interval). A lower number indicates a better score.

4.1. User Feedback/Impressions

Following each test we collected feedback and the comments are summarized below:

Visual Representation – The reference video on the screen (see figure 6) is rendered from the perspective of the camera, so the image is reversed from a "mirror". Most users preferred to see a mirror image of their body. We are planning to implement this transformation.

Ambidexterity – The initial system focuses on motion of a single part (joint) in the body. Most users were not equally as good at reproducing the teacher's path with both hands, particularly the spatial aspect.

Tolerance – Reproducing a path given the accuracy of the hardware and the user required us to have a configurable tolerance, effectively converting the path from a line to a "tube". For coarse full-body movements, several centimeters might be acceptable, but for more fine-grain hand motion, smaller tolerances may be used.

Sonification (Spatial) – Users all agree that the pitch change sonification helped them to know when and where they went off track, but it was not always clear which direction to move to correct. Currently we "bend" the pitch up linearly for deviations from the curve. We are investigating other mappings, which can use direction movement for pitch up/down. Sometimes the visual represents confused this further and better results were obtained from a user when the visual aids were turned off.

Sonification (Temporal) – Initial testing impressions of the ADSR envelope method were favorable. In particular, playback of the teacher's envelope prior to each exercise enabled them to develop a mental image of the sound to make when the correct timing is achieved.

Scoring – Our initial attempt was to simplifying scoring including both spatial and timing performance in one number. It became quickly apparent that we need to break this up into multiple components, so that the student can understand what needs to be improved, either timing or trajectory.

5. CONCLUSION AND FUTURE WORK

Our basic testing of Sound*Tracer* as an experimental platform for motion sonification and learning has yielded some initial results that are very encouraging. We have created a usable system that can map real-time movement to layers of sound, which can be used as a feedback loop for learning complex path-based motion. We are looking forward to further testing with more complex motion paths and sound mappings. In this initial effort we have focused on comprehensive path animation of a single point on the body over time with informal testing. As a next step, we can explore how to scale this to a fully articulated hierarchical skeleton. This could involve capturing the motion path for each joint or limb and comparing those to a reference set or perhaps investigating combining these methods with a pose-based approach. At a smaller physical scale, capturing more detailed skeletal features such as the hand could allow us to capture finger/hand based motion paths. Our method could prove to be a powerful technique of augmenting a gesture-based vocabulary with motion path information that includes both timing and spatial information.

6. DOCUMENTATION AND ACKNOWLEDGMENT

Video demonstrating the described system can be found at: http://www.youtube.com/user/k2msmith

We would like to thank the faculty, staff and students of California State University, Channel Islands for their support and feedback for the project. We would also like to

acknowledge the authors and project developers of the *Processing Language* and the *SuperCollider Real-Time Audio Synthesis and Algorithmic Composition* system, which are the key tools used in the development of the framework.

7. REFERENCES

[1] http://www.xbox.com/en-US/kinect
[2] http://www.leapmotion.com
[3] Niklas Rober and Maic Masuch, "Interacting with Sound, An Interaction Paradigm for Virtual Worlds", *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, Sidney Australia, July 2004.
[4] Alfred O. Effenberg, "Using Sonification to Enhance Perception and Reproduction Accuracy of Human Movement Patterns", *Proceedings of the Int. Workshop on Interactive Sonification*, Bielefeld, Germany, Jan 2004.
[5] Max Kleiman-Weiner and Jonathan Berger, "The Sound of One Arm Swinging: A Model for Multidimensional Auditory Display of Physical Motion", *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006),* London, UK, June 2006.
[6] Sanda Pauletto and Andy Hunt, "The Sonification of EMG Data", *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006),* London, UK, June 2006.
[7] Thomas Hermann, Andy Hunt, John G. Neuhoff, *The Sonification Handbook*, Logos Verlag, Berlin, Germany 2011.
[8] Katherina Vogt, David Pirro, Ingo Kobenz, Robert Holdrich and Gerhard Eckel, "PhysioSonic – Movement Sonification as Auditory Feedback, *Proceedings of the 15th International Conference on Auditory Display (ICAD 2009),* Copenhagen, Denmark, May 2009.
[9] http://www.synapsekinect.tumblr.com
[10] http://www.v.co.nz/ - the-motion-project
[11] http://ecmc.rochester.edu/ecmc/docs/supercollider/scbook/
[12] Perry R. Cook, *Real Sound Synthesis for Interactive Applications*, A K Peters/CRC Press, July 1, 2002.
[13] http://www.processing.org
[14] http://www.openni.org
[15] Scott Wilson, David Cottle and Nick Collins (edited), *The SuperCollider Book*, The MIT Press, Cambridge, Mass, 2011.
[16] http://doc.sccode.org
[17] http://www.sojamo.de/libraries/oscP5/
[18] Adrian Freed, Andy Schmeder, "Features and Future of Open Sound Control version 1.1 for NIME", *New Interfaces for Musical Expression (Conf) 2009*, Pittsburgh, PA, June 2009.
[19] http://hexler.net/software/touchosc